On the Pilot-Data Power Trade Off in Single Input Multiple Output Systems

Gábor Fodor*‡, Miklós Telek†

* Ericsson Research, Stockholm. Email: gabor.fodor@ericsson.com

‡ Royal Institute of Technology, Stockholm.

† Budapest University of Technology and Economics, Budapest. Email: telek@hit.bme.hu

***Abstract*—We consider a single cell single input multiple output (SIMO) system employing orthogonal frequency division multiplexing (OFDM). In such systems, setting the pilot-to-data power ratio (PDPR) has a large impact on the spectral and energy efficiency. In this paper we provide a closed form solution for the mean square error (MSE) of the received data as a function of the PDPR assuming Gaussian channels and minimum mean square error (MMSE) equalization. In numerical experiments we find that the MSE is a convex function of the PDPR and study the PDPR that minimizes the MSE as the number of antennas at the base station (BS) grows large. We find that the MSE-minimizing PDPR heavily depends on the number of antennas and the path loss between the mobile station and the BS. Specifically, as the number of antennas grows large, a larger portion of the total power budget needs to be allocated for pilot signals, especially for low path loss users.** [1]

***Index Terms*—cellular networks, MIMO systems, channel estimation, power control**

I. Introduction

In multiple input multiple output (MIMO) orthogonal frequency division multiplexing (OFDM) systems pilot symbols facilitate channel estimation, but they reduce the transmitted energy for data symbols under a fixed power budget. This fundamental tradeoff has been studied in [1] and [2] assuming various pilot patterns and receiver structures. An important insight from these works is that although balancing the pilot-to-data power ratio (PDPR) improves the system performance, the capacity is not highly sensitive to the PDPR over a fairly broad range. The results of [2], for example, indicate that the optimal PDPR provides about 2-3 dB gain compared with the equal power for pilot and data symbols.

Subsequently, [3] derived a closed form of the optimal PDPR in MIMO-OFDM spatial multiplexing systems with minimum mean square error channel estimation and showed that a tight bound lying in the quasi-optimal region provides a good approximation for the optimal PDPR setting. More recently, [4] derived a closed form PDPR that maximizes the capacity bound of MIMO-OFDM systems and studied the impact of carrier frequency offset (CFO) on the maximizing power allocation.

[1]The work of G. Fodor has been partially performed in the framework of the FP7 project ITC 317669 METIS. G. Fodor has also been supported by the Swedish Foundation for Strategic Research Strategic Mobility SM13-0008 Matthew Project. The work of M. Telek was partially performed when he was visiting the Royal Institute of Technology, Stockholm. M. Telek was partially supported by the OTKA K101150 project.

Along another line, results on massive MIMO systems suggest that as the number of base stations antennas grows large, the signal-to-noise-ratio becomes independent of the transmitted powers and the required transmitted energy per bit (i.e. for a given SNR target) vanishes [5]. This observation raises the question how the optimal PDPR is affected as the number of antennas at the base station grows large.

The direct precursor of the current paper is our previous work [6], in which we considered both a single cell and a multicell model, without calculating the mean square error for an arbitrary number of receive antennas. In this paper we consider a single input multiple output (SIMO) system in which the mobile station balances its PDPR, while the base station uses maximum likelihood (ML) channel estimation to initialize a linear minimum mean square error (MMSE) equalizer. Our contribution is to derive a closed form for the mean square error (MSE) of the equalized data symbols. This form is powerful, because it includes not only the pilot and data transmit power levels as independent variables, but also the number of receive antennas at the base station (N_r). This formula allows us to study the impact of N_r on the MSE and thereby on the PDPR that minimizes the MSE. To the best of our knowledge the MSE formula as well as the insights obtainied in the numerical section are novel.

We organize the rest of this paper as follows. The system model is defined in Section II. The MSE is determined in Section III. Numerical results are studied in Section IV. Section V concludes the paper.

II. System Model

We consider the uplink transmission of a SIMO single cell multi-user wireless system, in which users are scheduled on orthogonal frequency channels. It is assumed that each mobile station (MS) employs an orthogonal pilot sequence, so that no interference between pilots is present in the system. This is a common assumption in massive multi-user MIMO systems in which a single MS may have a single antenna [5]. Since the channel is quasi-static frequency-flat within each transmission block, it is equivalent to model the whole pilot sequence as a single symbol per resource block with power P^p , while each data symbol is transmitted with power P. The base station (BS) estimates the channel $\mathbf{h}$ (column vector of dimension N_r, where N_r is the number of receive antennas at the BS) by employing maximum likelihood (ML)

channel estimators to initialize linear minimum mean square error (MMSE) equalizers.

A. Channel Estimation Model

Each MS transmits an orthogonal pilot symbol x_j that is received by the BS. Thus, the column vector of the received pilot signal at the BS from the j^{th} MS is:

$$\mathbf{y}_j^p = \sqrt{P_j^p}\alpha_j \mathbf{h}_j x_j + \mathbf{n}^p, \tag{1}$$

where we assume that $\mathbf{h}_j$ is a circular symmetric complex normal distributed vector of r.v. with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{C}_j$ (of size N_r), denoted as $\mathbf{h}_j \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_j)$, α_j accounts for the propagation loss, $\mathbf{n}^p \sim \mathcal{CN}(\mathbf{0}, \sigma^2\ \mathbf{I})$ is the contribution from additive Gaussian noise and the pilot symbol is scaled as $|x_j|^2 = 1, \forall j$. Since we assume orthogonal pilot sequences, the channel estimation process can be assumed independent for each MS and we can therefore drop the index j. With a ML channel estimator, the BS estimates the channel based on (1) assuming

$$\hat{\mathbf{h}} = \frac{\mathbf{y}^p}{\sqrt{P^p}\alpha x},$$

that is:

$$\hat{\mathbf{h}} = \mathbf{h} + \frac{\mathbf{n}^p}{\sqrt{P^p}\alpha x}; \quad |x|^2 = 1. \tag{2}$$

It then follows that the estimated channel $\hat{\mathbf{h}}$ is distributed as follows:

$$\hat{\mathbf{h}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}), \tag{3}$$

with $\mathbf{R} \triangleq \mathcal{E}[\hat{\mathbf{h}}\hat{\mathbf{h}}^H] = \mathbf{C} + \frac{\sigma^2}{P^p\alpha^2}\ \mathbf{I}$.

Further, it follows that the channel estimation error $\mathbf{w} \triangleq \hat{\mathbf{h}} - \mathbf{h}$ is also normally distributed with a covariance inversely proportional to the employed pilot power:

$$\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C_w}); \quad \mathbf{C_w} \triangleq \frac{\sigma^2}{P^p\alpha^2}\ \mathbf{I}.$$

Equations (2)-(3) imply that $\mathbf{h}$ and $\hat{\mathbf{h}}$ are jointly circular symmetric complex Gaussian (multivariate normal) distributed random variables [7], [8]. Specifically, we recall from [7] that the covariance matrix of the joint PDF is composed by autocovariance matrices $\mathbf{C_{h,h}}$, $\mathbf{C_{\hat{h},\hat{h}}}$ and cross covariance matrices $\mathbf{C_{h,\hat{h}}}$, $\mathbf{C_{\hat{h},h}}$ as

$$\begin{bmatrix} \mathbf{C_{h,h}} & \mathbf{C_{h,\hat{h}}} \\ \mathbf{C_{\hat{h},h}} & \mathbf{C_{\hat{h},\hat{h}}} \end{bmatrix} = \begin{bmatrix} \mathbf{C} & \mathbf{C} \\ \mathbf{C} & \mathbf{R} \end{bmatrix},$$

and $\mathbf{R} = \mathbf{C} + \mathbf{C_w}$.

B. Determining the Conditional Channel Distribution

From the joint PDF of $\mathbf{h}$ and $\hat{\mathbf{h}}$ we can compute the following conditional distributions.

Result 2.1: Given a random channel realization $\mathbf{h}$, the estimated channel $\hat{\mathbf{h}}$ conditioned to $\mathbf{h}$ can be shown to be distributed as

$$(\hat{\mathbf{h}} \mid \mathbf{h}) \sim \mathbf{h} + \mathcal{CN}(\mathbf{0}, \mathbf{C_w}). \tag{4}$$

Result 2.2: The distribution of the channel realization $\mathbf{h}$ conditioned to the estimate $\hat{\mathbf{h}}$ is normally distributed as follows:

$$(\mathbf{h} \mid \hat{\mathbf{h}}) \sim \mathbf{D}\hat{\mathbf{h}} + \mathcal{CN}\Big(\mathbf{0}, \mathbf{Q}\Big), \tag{5}$$

where $\mathbf{D} = \mathbf{C}\mathbf{R}^{-1}$ and $\mathbf{Q} = \mathbf{C} - \mathbf{C}\mathbf{R}^{-1}\mathbf{C}$.
The proofs are provided in the Appendix.

To capture the tradeoff between the pilot and data power, we need to calculate the mean square error of the equalized data symbols. To this end, we consider an equalization model in the next subsection.

C. Equalizer Model based on ML Channel Estimator

The data signal received by the BS is

$$\mathbf{y} = \alpha\sqrt{P}\mathbf{h}x + \mathbf{n}, \tag{6}$$

where $|x|^2 = 1$. We assume that the BS employs a naive MMSE equalizer, where the estimated channel (2) is taken as if it was the actual channel:

$$\mathbf{G} = \alpha\sqrt{P}\hat{\mathbf{h}}^H(\alpha^2 P\hat{\mathbf{h}}\hat{\mathbf{h}}^H + \sigma^2\mathbf{I})^{-1}. \tag{7}$$

Under this assumption, we state the following result as a first step towards determining the MSE.

Result 2.3: Let $\text{MSE}(\mathbf{h}, \hat{\mathbf{h}}) = \mathcal{E}_{x,\mathbf{n}}\left\{|\mathbf{G}\mathbf{y} - x|^2\right\}$ be the MSE for the equalized symbols, given the realizations of $\mathbf{h}$ and $\hat{\mathbf{h}}$. It is

$$\begin{aligned}\text{MSE}(\mathbf{h}, \hat{\mathbf{h}}) = \alpha^2 P\mathbf{G}\mathbf{h}\mathbf{h}^H\mathbf{G}^H - \\ - 2\alpha\sqrt{P}Re[\mathbf{G}\mathbf{h}] + \sigma^2\mathbf{G}\mathbf{G}^H + 1.\end{aligned} \tag{8}$$

The proof is presented in the Appendix. From this, our next result follows directly.

Result 2.4: Let $\text{MSE}(\hat{\mathbf{h}}) = \mathcal{E}_{\mathbf{h}|\hat{\mathbf{h}}}\Big\{\text{MSE}(\mathbf{h}, \hat{\mathbf{h}})\Big\}$ be the MSE for the equalized symbols, given the estimated channel realization $\hat{\mathbf{h}}$. It satisfies

$$\begin{aligned}\text{MSE}(\hat{\mathbf{h}}) = \mathbf{G}\Big(\alpha^2 P(\mathbf{D}\hat{\mathbf{h}}\hat{\mathbf{h}}^H\mathbf{D}^H + \mathbf{Q}) + \sigma^2\mathbf{I}\Big)\mathbf{G}^H - \\ - 2\alpha\sqrt{P}Re\{\mathbf{G}\mathbf{D}\hat{\mathbf{h}}\} + 1.\end{aligned} \tag{9}$$

The proof is presented in the Appendix.

III. Determining the Unconditional MSE

Based on the conditional MSE expression of the preceding section, we are now interested in deriving the unconditional expectation of the MSE. To this end, the following two lemmas turn out to be useful.

Lemma *1:* Given a channel estimate instance $\hat{\mathbf{h}}$, the MMSE weighting matrix $\mathbf{G}$, as a function of the number of receive antennas at the base station (N_r) can be expressed as follows

$$\mathbf{G} = \frac{\alpha\sqrt{P}}{\|\hat{\mathbf{h}}\|^2\alpha^2 P + \sigma^2}\ \hat{\mathbf{h}}^H, \tag{10}$$

where $\|\hat{\mathbf{h}}\|^2 = \hat{\mathbf{h}}^H\hat{\mathbf{h}} = \sum_{i=1}^{N_r}|\hat{h}_i|^2$.
The proof is presented in the Appendix.

Using this simple expression of $\mathbf{G}$ we can further simplify the conditional expectation of the MSE of the MMSE equalized data symbols.

Lemma *2:* When we assume independent channel distributions with identical variances, that is the channel covariance matrix is diagonal in the form of $\mathbf{C} = \varrho\mathbf{I}$, where $\varrho \in \mathbb{R}^+$, then the covariance matrices $\mathbf{D}$, $\mathbf{Q}$ are $\mathbf{D} = d\mathbf{I}$, $\mathbf{Q} = q\mathbf{I}$ with $d = \varrho(\varrho + \frac{\sigma^2}{P^p\alpha^2})^{-1}$, $q = \varrho(1-d)$ and (9) simplifies to

$$\mathrm{MSE}(\hat{\mathbf{h}}) = 1 - \frac{2\|\hat{\mathbf{h}}\|^2 d\alpha^2 P}{\|\hat{\mathbf{h}}\|^2\alpha^2 P + \sigma^2} + \\ + \frac{\|\hat{\mathbf{h}}\|^2\alpha^2 P}{\left(\|\hat{\mathbf{h}}\|^2\alpha^2 P + \sigma^2\right)^2} \cdot \left[\|\hat{\mathbf{h}}\|^2 d^2\alpha^2 P + q\alpha^2 P + \sigma^2\right] \quad (11)$$

The proof is presented in the Appendix. After these preparations we can state the main theorem about the MSE.

Theorem *1:* The expected value of the mean square error of the equalized symbols is

$$\mathcal{E}\{MSE\} = d^2 N_r \Bigg(\mathcal{G}(a, 1+N_r) + \\ + pr\mathcal{G}(1+N_r, 1+N_r) - 1 \Bigg) + \\ + \frac{b}{pr}\left(\mathcal{G}(a, N_r) + pr\mathcal{G}(N_r, N_r) - 1 \right) - \\ - 2d \cdot \left(pr\mathcal{G}(N_r, 1+N_r) \right) + 1;$$

where $p = \alpha^2 P$,

$$\mathcal{G}(x,y) \triangleq \frac{1}{pr} e^{\frac{a}{pr}} x E_{in}\left(y, \frac{a}{pr}\right),$$

and $E_{in}(n,z) \triangleq \int_1^\infty e^{-zt}/t^n \; dt$ is a standard exponential integral function which is commonly available in numerical programming environments (e.g., it is called ExpIntegralE in Mathematica).

The proof is provided in the Appendix.

It is important to note that q and d carry the dependency on the pilot power P^p and p carries the dependency on the data power P in $\mathcal{E}\{MSE\}$.

IV. Numerical Results

In this section we consider a single cell SIMO system and concentrate on the performance of a single mobile station (MS) scheduled on a flat fading frequency channel. Unless stated otherwise, we assume that the MS has a power budget of 24 dBm that needs to be shared between the pilot and data symbols, as described in Section II. Except for Figures 9-11, the MS is in a position within the cell at which the distance dependent path loss is 60 dB.

Figure 1-2 are contour plots of the MSE of the equalized symbols as a function of the employed pilot and data transmit power levels when the number of receive antennas is $N_r = 2$ and $N_r = 100$ respectively. These figures indicate the pilot-data transmit power level pairs that maintain a given MSE. For example, in Figure 1 we can see that the lowest MSE value that is feasible with a 250 mW power budget is 0.5. In contrast, Figure 2 shows that when $N_r = 100$, the same power budget can maintain an MSE less than 0.1. From this figure it is also clear that the 'knee' of the MSE curves is shifted toward much lower data power levels, which intuitively suggests a shift in the optimal PDPR. For example, the optimal MSE with 250 mW power budget is attained at around $P^p = 145, P = 105$ on Figure 1 and around $P^p = 215, P = 35$ on Figure 2.

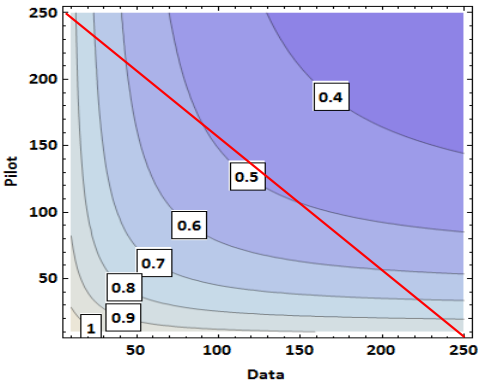

Fig. 1. Contour plot of the MSE achieved by specific pilot and data power settings of a SIMO system with $N_r = 2$ receiver antennas. The diagonal line indicates the feasible region of a mobile station of a sum power level of 250 mW.

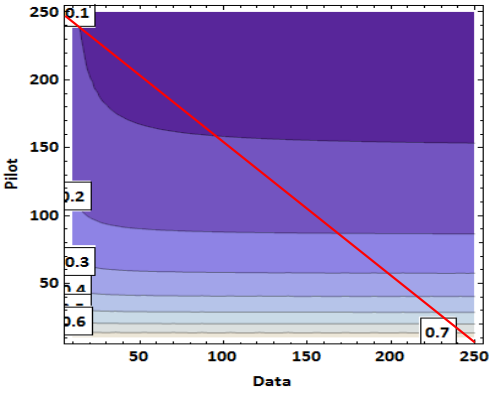

Fig. 2. Contour plot of the MSE achieved by specific pilot and data power settings of a SIMO system with 100 receiver antennas. The diagonal line indicates the feasible region of a mobile station of a sum power level of 250 mW.

Figure 3 shows the impact of increasing the number of antennas at the base station from 2 to 100 in terms of the MSE performance as the function of the pilot and data transmit power. On the lower plane ($N_r = 100$), the pilot power minimizing the MSE is shifted towards a higher value compared with the $N_r = 2$ case (indicated with a circle) when we assume a power budget of 250 mW.

For illustration purposes, Figures 4, 5 and 6 examine the convexity of the MSE as a function of the pilot and data power levels. Figure 4 considers the case of a fixed sum power budget (250 mW) and plots the MSE as a function of the data power

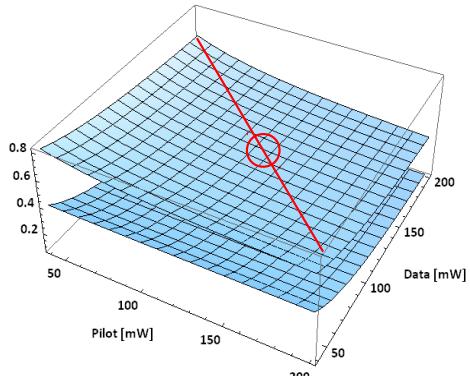

Fig. 3. The MSE of a SIMO system of 2 and 100 antennas. The circle indicates the optimal pilot and data power setting for the 2 antenna system with a sum power constraint of 250 mW.

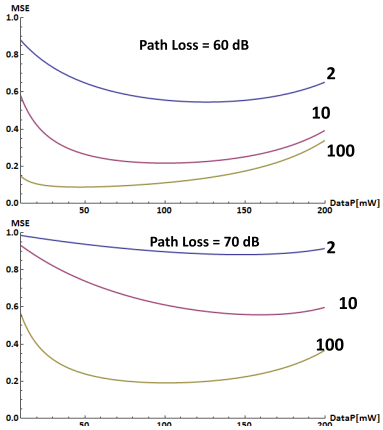

Fig. 4. The MSE as a function of the data power at path loss 60 dB and 70 dB with 2, 10 and 100 receive antennas at the base station. At 60 dB path loss, for example, the data power that minimizes the MSE is P=105 mw (2 antennas) and P=35 mW (100 antennas).

with different path loss values. In this numerical example the MSE is clearly a convex function of the data power. Figures 5-6 illustrate the eigenvalues of the Hessian matrix of the MSE as a function of the data and pilot power levels with $N_r = 2$ and $N_r = 100$ respectively. These figures illustrate that the Hessian matrix of the MSE is positive definite implying the convexity of the MSE. We note that since the MSE in general is a function of the geometry of the system and the number of antennas in addition to the transmit power levels, the symbolic examination of its convexity becomes intractable.

These results are reinforced by Figure 7 that shows the MSE as a function of the allocated pilot power under varying power budget (200mW, 225 mW and 250 mW) and assuming different number of receive antennas ($N_r = 2$, $N_r = 20$ and $N_r = 100$). Here we can clearly see the tendency that as the number of the antennas grows large, the MS needs to allocate a smaller share of the total budget to data transmission and can 'afford' a larger share of the budget for pilot transmission.

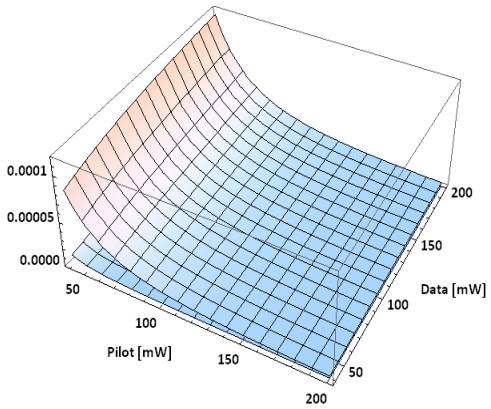

Fig. 5. The eigenvalues of the Hessian matrix of the MSE as a function of the data and pilot power levels with $N_r = 2$ receive antennas. Both eigenvalues are strictly positive.

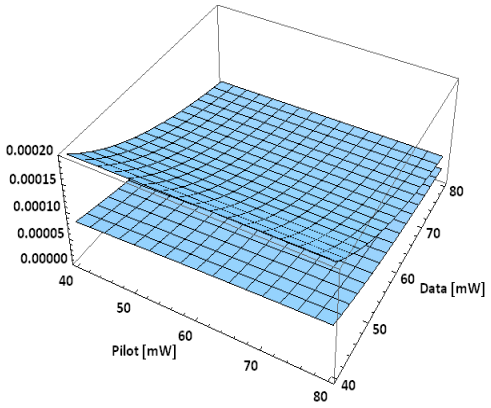

Fig. 6. The eigenvalues of the Hessian matrix of the MSE as a function of the data and pilot power levels with $N_r = 100$ receive antennas. Both eigenvalues are strictly positive.

This basic insight is in line with the classical observation by Marzetta predicting a diminishing data transmit power required for maintaining an SNR target [5].

Figure 8 examines how the required pilot and data power levels that are necessary to maintain a predefined MSE level depend on that predefined MSE value when the number of antennas grows. As we can expect, larger number of antennas means that the pilot-data power balance becomes less sensitive to the MSE target.

Figure 9 shows the MSE as a function of the data power and the path loss for two antenna configurations ($N_r = 2$ and $N_r = 100$). We can observe that the data power level that minimizes the MSE is not only dependent on N_r, but also on the path loss. Specifically, for larger path loss (cell edge) users, more data power (i.e. less pilot power) minimizes the MSE than for cell center users. However, this effect becomes less pronounced as the number of antennas increases.

This basic observation is further examined by Figure 10 and Figure 11. Figure 10 plots the ratio of the power budget that

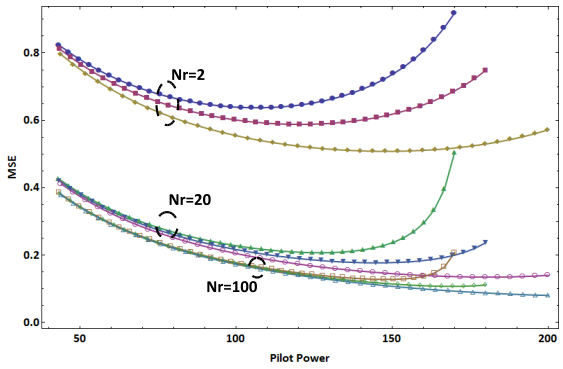

Fig. 7. The MSE as a function of the pilot power of a SIMO system with $N_r = 2, 20, 100$ antennas respectively, for 3 different sum power constraints (200 mW, 225 mW and 250 mW). As the number of antennas increases, the optimal pilot power increases.

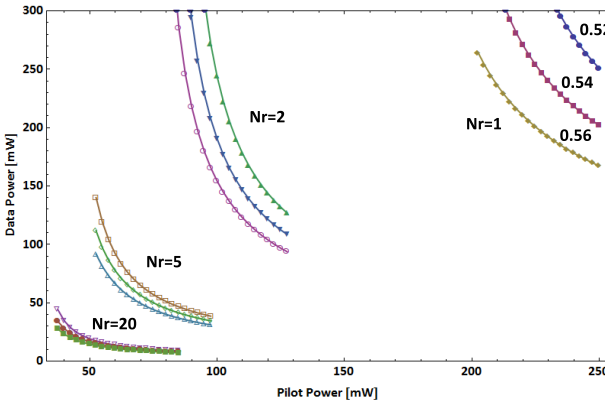

Fig. 8. Pilot and data power levels that are necessary to maintain predefined MSE targets (0.52, 0.54, 0.56) in a SIMO system with 1, 2, 5 and 20 antennas at the BS. The required pilot-data setting is less sensitive to small changes in the MSE target with increasing number of antennas.

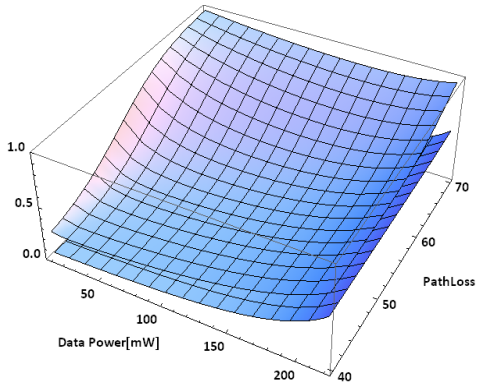

Fig. 9. The MSE as a function of the data power and the distance dependent path loss of a sum power constrained (250 mW) SIMO system with $N_r = 2$ and $N_r = 100$ antennas. For $N_r = 2$, as the path loss increases, the data power level that minimizes the MSE increases. However, this effect is not visible for $N_r = 100$.

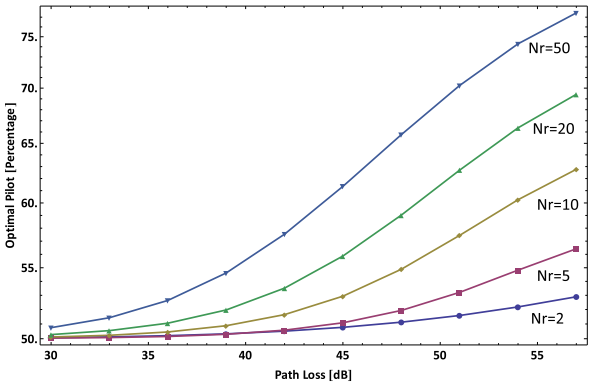

Fig. 10. The ratio of the pilot power that minimizes the MSE when the total power budget is set to 250 mW, as a function of the path loss, for 5 different antenna configurations (2, 5, 10, 20 and 50 antennas). As the path loss increases, a larger portion of the power budget needs to be spent on the pilot power, especially with a greater number of antennas.

needs to be spent on the plot signal such that the MSE is minimized for $N_r = 2, 5, 10, 20, 50$. With a greater number of antennas, the optimal pilot setting becomes more sensitive to the path loss between the MS and the BS. For example, with 20 antennas at the BS, around 70% of the power budget must be spent on the pilot signal for a MS experiencing 57 dB path loss, while for a MS at around 30 dB, the MS needs to spend 50% of the power budget on pilot signaling virtually independently of the number of antennas.

On the other hand, as shown in Figure 11, when the pilot-to-data power ratio is set optimally, the resulting minimum MSE is orders of magnitude lower when the number of antennas grows large. For example, with $N_r = 2$ antennas, the minimal MSE grows from around 0.08 to 0.44 as the path loss increases from 30 dB to 57 dB, while with $N_r = 50$ antennas, this increase is from around $5 \cdot 10^{-5}$ to $7 \cdot 10^{-2}$.

V. Concluding Remarks

The main contribution of this paper is the derivation of the MSE as the function of the employed pilot and data power levels as well as the number of receive antennas in SIMO OFDM systems. The numerical results provide two key insights. First, as the number of antennas at the base station increases, the MSE is minimized when a larger portion of the total transmit power budget is allocated for pilot transmission. This result is in line with the results from massive MIMO systems that suggest that the required transmit energy per bit vanishes as the number of antennas grows large. Secondly, as the path loss between the MS transmitting the pilot and base station increases, a smaller portion of the power budget needs to be spent on the pilot power. This second effect becomes less pronounced as the number of antennas at the base station increases. Our summary is therefore that the PDPR that minimizes the MSE of the equalized symbols heavily depends on both the number of antennas and the MS position within the cell. An important future work is to investigate multicell systems, in which greater pilot power does not only imply

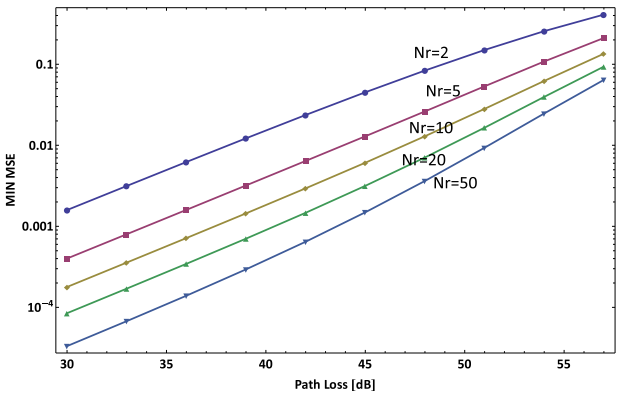

Fig. 11. The minimum MSE value that can be reached by optimal pilot power setting, as a function of the path loss, for 5 different antenna configurations (2,5,10,20 and 50 antennas). As the number of antennas increases, a close to zero MSE can be ensured for a much greater range of path loss values than with a few antennas.

lower available power for data transmission, but also a higher level of pilot contamination [5]. Therefore, our conclusions from the single cell analysis need to be reexamined in multicell systems.

ACKNOWLEDGMENTS

M. Telek was partially supported by the OTKA K101150 project.

APPENDIX

Proof of Result 2.1: To prove the result we apply (10.24)-(10.28) of [7], but in contrast with [7], (4) and (5), in this proof we explicitly distinguish between the condition, $\mathbf{h_0}$, and the unconditional random vector, $\mathbf{h}$. According to [7] the conditional distribution of $\hat{\mathbf{h}}|\mathbf{h_0}$ is complex normal with the following properties:

$$E(\hat{\mathbf{h}}|\mathbf{h_0}) = \underbrace{E(\hat{\mathbf{h}})}_{\mathbf{0}} + \underbrace{\mathbf{C}_{\hat{\mathbf{h}},\mathbf{h}}}_{\mathbf{C}} \underbrace{\mathbf{C}_{\mathbf{h},\mathbf{h}}^{-1}}_{\mathbf{C}^{-1}} \left(\mathbf{h_0} - \underbrace{E(\mathbf{h})}_{\mathbf{0}}\right) = \mathbf{h_0}.$$

$$\mathbf{C}_{\hat{\mathbf{h}}|\mathbf{h_0}} = \mathbf{C}_{\hat{\mathbf{h}},\hat{\mathbf{h}}} - \mathbf{C}_{\hat{\mathbf{h}},\mathbf{h}}\mathbf{C}_{\mathbf{h},\mathbf{h}}^{-1}\mathbf{C}_{\mathbf{h},\hat{\mathbf{h}}} = \mathbf{R} - \mathbf{C} = \mathbf{C}_{\mathbf{w}};$$

■

Proof of Result 2.2: Similarly to the proof of Result 2.1, $\mathbf{h}|\hat{\mathbf{h}}_\mathbf{0}$ is complex normal distributed with the following mean and covariance [7]

$$E(\mathbf{h}|\hat{\mathbf{h}}_\mathbf{0}) = E(\mathbf{h}) + \mathbf{C}_{\mathbf{h},\hat{\mathbf{h}}}\mathbf{C}_{\hat{\mathbf{h}},\hat{\mathbf{h}}}^{-1}\left(\hat{\mathbf{h}}_\mathbf{0} - E(\hat{\mathbf{h}})\right) = \mathbf{C}\mathbf{R}^{-1}\hat{\mathbf{h}}_\mathbf{0};$$

$$\mathbf{C}_{\mathbf{h}|\hat{\mathbf{h}}_\mathbf{0}} = \mathbf{C}_{\mathbf{h},\mathbf{h}} - \mathbf{C}_{\mathbf{h},\hat{\mathbf{h}}}\mathbf{C}_{\hat{\mathbf{h}},\hat{\mathbf{h}}}^{-1}\mathbf{C}_{\hat{\mathbf{h}},\mathbf{h}} = \mathbf{C} - \mathbf{C}\mathbf{R}^{-1}\mathbf{C}.$$

■

Proof of Result 2.3: Having $\mathbf{y} = \alpha\sqrt{P}\mathbf{h}x + \mathbf{n}$ the mean square error of the equalized symbols, given a specific set of realizations $\mathbf{h}$ and $\hat{\mathbf{h}}$ can be calculated as:

$$\begin{aligned}\text{MSE}\left(\mathbf{h},\hat{\mathbf{h}}\right) &= \mathcal{E}_{x,\mathbf{n}}\left\{|\mathbf{G}\mathbf{y} - x|^2\right\} = \\ &= \mathcal{E}_{x,\mathbf{n}}\left\{|\underbrace{(\mathbf{G}\alpha\sqrt{P}\mathbf{h} - 1)x}_{a} + \underbrace{\mathbf{G}\mathbf{n}}_{b}|^2\right\} \end{aligned} \quad (12)$$

using $|a+b|^2 = (a+b)(a^H+b^H) = aa^H + ab^H + a^Hb + bb^H$ we further have

$$\begin{aligned}&\text{MSE}\left(\mathbf{h},\hat{\mathbf{h}}\right) = \\ &= \mathcal{E}_{x,\mathbf{n}}\left\{((\mathbf{G}\alpha\sqrt{P}\mathbf{h} - 1)x)((\mathbf{G}\alpha\sqrt{P}\mathbf{h} - 1)x)^H\right\} + \\ &\quad + \mathcal{E}_{x,\mathbf{n}}\left\{((\mathbf{G}\alpha\sqrt{P}\mathbf{h} - 1)x)(\mathbf{G}\mathbf{n})^H\right\} + \\ &\quad + \mathcal{E}_{x,\mathbf{n}}\left\{\mathbf{G}\mathbf{n}((\mathbf{G}\alpha\sqrt{P}\mathbf{h} - 1)x)^H\right\} + \\ &\quad + \mathcal{E}_{x,\mathbf{n}}\left\{\mathbf{G}\mathbf{n}(\mathbf{G}\mathbf{n})^H\right\} = \\ &= (\mathbf{G}\alpha\sqrt{P}\mathbf{h} - 1)\underbrace{\mathcal{E}_{x,\mathbf{n}}\{xx^H\}}_{1}(\mathbf{G}\alpha\sqrt{P}\mathbf{h} - 1)^H + \\ &\quad + (\mathbf{G}\alpha\sqrt{P}\mathbf{h} - 1)\underbrace{\mathcal{E}_{x,\mathbf{n}}\{x\}}_{0}\mathcal{E}_{x,\mathbf{n}}\{\mathbf{n}^H\}\mathbf{G}^H + \\ &\quad + \mathbf{G}\mathcal{E}_{x,\mathbf{n}}\{\mathbf{n}\}\underbrace{\mathcal{E}_{x,\mathbf{n}}\{x\}}_{0}(\mathbf{G}\alpha\sqrt{P}\mathbf{h} - 1)^H + \\ &\quad + \mathbf{G}\underbrace{\mathcal{E}_{x,\mathbf{n}}\{\mathbf{n}\mathbf{n}^H\}}_{\sigma^2\mathbf{I}}\mathbf{G}^H = \\ &= \left|\mathbf{G}\mathbf{h}\alpha\sqrt{P} - 1\right|^2 + \sigma^2\mathbf{G}\mathbf{G}^H\end{aligned}$$

and finally

$$\begin{aligned}&\text{MSE}\left(\mathbf{h},\hat{\mathbf{h}}\right) = \\ &= \left|\mathbf{G}\mathbf{h}\alpha\sqrt{P} - 1\right|^2 + \sigma^2\mathbf{G}\mathbf{G}^H = \\ &= \left(\mathbf{G}\mathbf{h}\alpha\sqrt{P} - 1\right)\cdot\left(\mathbf{G}\mathbf{h}\alpha\sqrt{P} - 1\right)^H + \sigma^2\mathbf{G}\mathbf{G}^H = \\ &= \left(\mathbf{G}\mathbf{h}\alpha\sqrt{P} - 1\right)\cdot\left(\mathbf{h}^H\mathbf{G}^H\alpha\sqrt{P} - 1\right) + \sigma^2\mathbf{G}\mathbf{G}^H = \\ &= \alpha^2P\mathbf{G}\mathbf{h}\mathbf{h}^H\mathbf{G}^H - \mathbf{G}\mathbf{h}\alpha\sqrt{P} - \mathbf{h}^H\mathbf{G}^H\alpha\sqrt{P} + \\ &\quad + 1 + \sigma^2\mathbf{G}\mathbf{G}^H = \\ &= \alpha^2P\mathbf{G}\mathbf{h}\mathbf{h}^H\mathbf{G}^H - \alpha\sqrt{P}\left(\mathbf{G}\mathbf{h} + \mathbf{h}^H\mathbf{G}^H\right) + \\ &\quad + 1 + \sigma^2\mathbf{G}\mathbf{G}^H = \\ &= \alpha^2P\mathbf{G}\mathbf{h}\mathbf{h}^H\mathbf{G}^H - \alpha\sqrt{P}\ 2Re\left[\mathbf{G}\mathbf{h}\right] + \\ &\quad + 1 + \sigma^2\mathbf{G}\mathbf{G}^H.\end{aligned}$$

■

Proof of Result 2.4: We can compute $\mathcal{E}_{\mathbf{h}|\hat{\mathbf{h}}}\left\{\text{MSE}\left(\mathbf{h},\hat{\mathbf{h}}\right)\right\}$ based on the conditional distribution given in (5):

$$\left(\mathbf{h}|\hat{\mathbf{h}}\right) \sim \mathbf{D}\hat{\mathbf{h}} + \mathcal{CN}(\mathbf{0},\mathbf{Q}).$$

Recall that for a complex random column vector $\mathbf{X}$:

$$\mathcal{E}(\mathbf{X}\mathbf{X}^H) = \mathcal{E}(\mathbf{X})\ \mathcal{E}(\mathbf{X})^H + Cov(\mathbf{X}).$$

Consequently,

$$\mathcal{E}\{\mathbf{h}|\hat{\mathbf{h}}\} = \mathbf{D}\hat{\mathbf{h}} \ \text{ and } \ \mathcal{E}\{\mathbf{h}\mathbf{h}^H|\hat{\mathbf{h}}\} = \mathbf{D}\hat{\mathbf{h}}\hat{\mathbf{h}}^H\mathbf{D}^H + \mathbf{Q}\ .$$

The expectation reads as:

$$\begin{aligned}
&\text{MSE}\left(\hat{\mathbf{h}}\right) = \mathcal{E}_{\mathbf{h}|\hat{\mathbf{h}}}\left\{\text{MSE}\left(\mathbf{h},\hat{\mathbf{h}}\right)\right\} \qquad (13)\\
&= \mathcal{E}_{\mathbf{h}|\hat{\mathbf{h}}}\left\{\mathbf{G}\mathbf{h}\mathbf{h}^H\mathbf{G}^H\alpha^2P\right\} - \mathcal{E}_{\mathbf{h}|\hat{\mathbf{h}}}\left\{2\alpha\sqrt{P}Re\big[\mathbf{G}\mathbf{h}\big]\right\} + \\
&\quad +\sigma^2\mathbf{G}\mathbf{G}^H + 1 = \\
&= \mathbf{G}\mathcal{E}_{\mathbf{h}|\hat{\mathbf{h}}}\left\{\mathbf{h}\mathbf{h}^H\right\}\mathbf{G}^H\alpha^2P - 2\alpha\sqrt{P}Re\big[\mathbf{G}\mathcal{E}_{\mathbf{h}|\hat{\mathbf{h}}}\left\{\mathbf{h}\right\}\big] + \\
&\quad +\sigma^2\mathbf{G}\mathbf{G}^H + 1 = \\
&= \mathbf{G}(\mathbf{D}\hat{\mathbf{h}}\hat{\mathbf{h}}^H\mathbf{D}^H + \mathbf{Q})\mathbf{G}^H\alpha^2P - 2\alpha\sqrt{P}Re\big[\mathbf{G}\mathbf{D}\hat{\mathbf{h}}\big] + \\
&\quad +\sigma^2\mathbf{G}\mathbf{G}^H + 1 \\
&= \mathbf{G}\left(\alpha^2P(\mathbf{D}\hat{\mathbf{h}}\hat{\mathbf{h}}^H\mathbf{D}^H + \mathbf{Q}) + \sigma^2\mathbf{I}\right)\mathbf{G}^H - \\
&\quad -2\alpha\sqrt{P}Re\big[\mathbf{G}\mathbf{D}\hat{\mathbf{h}}\big] + 1.
\end{aligned}$$

■

Proof of Lemma 1: According to the matrix inversion lemma for matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ of size $n\times n$, $n\times m$, $m\times m$, $m\times n$, respectively, we have

$$\begin{aligned}
&(\mathbf{A}+\mathbf{B}\mathbf{C}\mathbf{D})^{-1} = \\
&\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{D}\mathbf{A}^{-1}\mathbf{B} + \mathbf{C}^{-1}\right)^{-1}\mathbf{D}\mathbf{A}^{-1}\ .
\end{aligned}$$

Substituting $\mathbf{A} = \sigma^2\mathbf{I}$, $\mathbf{B} = \alpha\sqrt{P}\hat{\mathbf{h}}$, $\mathbf{C} = 1$, $\mathbf{D} = \alpha\sqrt{P}\hat{\mathbf{h}}^H$ we have

$$\begin{aligned}
&(\sigma^2\mathbf{I} + \alpha^2P\hat{\mathbf{h}}\hat{\mathbf{h}}^H)^{-1} = \frac{1}{\sigma^2}\mathbf{I} - \frac{1}{\sigma^2}\mathbf{I}\alpha\sqrt{P}\hat{\mathbf{h}}\cdot\\
&\cdot\left(\alpha\sqrt{P}\hat{\mathbf{h}}^H\frac{1}{\sigma^2}\mathbf{I}\alpha\sqrt{P}\hat{\mathbf{h}} + 1\right)^{-1}\alpha\sqrt{P}\hat{\mathbf{h}}^H\frac{1}{\sigma^2}\mathbf{I} = \\
&= \frac{1}{\sigma^2}\mathbf{I} - \frac{\frac{\alpha^2P}{\sigma^4}\hat{\mathbf{h}}\hat{\mathbf{h}}^H}{\frac{\alpha^2P}{\sigma^2}\hat{\mathbf{h}}^H\hat{\mathbf{h}} + 1},
\end{aligned}$$

where $\hat{\mathbf{h}}^H\hat{\mathbf{h}} = \|\hat{\mathbf{h}}\|^2$. Finally,

$$\begin{aligned}
\mathbf{G} &= \alpha\sqrt{P}\hat{\mathbf{h}}^H(\alpha^2P\hat{\mathbf{h}}\hat{\mathbf{h}}^H + \sigma^2\mathbf{I})^{-1}\\
&= \frac{\alpha\sqrt{P}}{\sigma^2}\hat{\mathbf{h}}^H - \frac{\alpha\sqrt{P}\frac{\alpha^2P}{\sigma^4}\hat{\mathbf{h}}^H\hat{\mathbf{h}}}{\frac{\alpha^2P}{\sigma^2}\|\hat{\mathbf{h}}\|^2 + 1}\hat{\mathbf{h}}^H\\
&= \frac{\alpha\sqrt{P}}{\sigma^2}\hat{\mathbf{h}}^H\left(1 - \frac{\alpha^2P\|\hat{\mathbf{h}}\|^2}{\alpha^2P\|\hat{\mathbf{h}}\|^2 + \sigma^2}\right)\\
&= \frac{\alpha\sqrt{P}}{\sigma^2}\hat{\mathbf{h}}^H\left(\frac{\sigma^2}{\alpha^2P\|\hat{\mathbf{h}}\|^2 + \sigma^2}\right)\\
&= \frac{\alpha\sqrt{P}}{\alpha^2P\|\hat{\mathbf{h}}\|^2 + \sigma^2}\hat{\mathbf{h}}^H.
\end{aligned}$$

■

Proof of Lemma 2: To simplify the notation we introduce z such that

$$\mathbf{G} = \frac{\alpha\sqrt{P}}{\|\hat{\mathbf{h}}\|^2\alpha^2P + \sigma^2}\ \hat{\mathbf{h}}^H = z\ \hat{\mathbf{h}}^H$$

Substituting this into (9) and using $\mathbf{D} = d\mathbf{I}$, $\mathbf{Q} = q\mathbf{I}$ yields (since $d\in\mathbb{R}^+$):

$$\begin{aligned}
&\text{MSE}\left(\hat{\mathbf{h}}\right) = \\
&= z\hat{\mathbf{h}}^H\left[\alpha^2P\left(d\mathbf{I}\hat{\mathbf{h}}\hat{\mathbf{h}}^Hd\mathbf{I} + q\mathbf{I}\right) + \sigma^2\mathbf{I}\right]z\hat{\mathbf{h}} -\\
&- 2\alpha\sqrt{P}\text{Re}\left\{z\hat{\mathbf{h}}^Hd\mathbf{I}\hat{\mathbf{h}}\right\} + 1 = \\
&= z\hat{\mathbf{h}}^H\left[\alpha^2P\left(d^2\hat{\mathbf{h}}\hat{\mathbf{h}}^H + q\mathbf{I}\right) + \sigma^2\mathbf{I}\right]z\hat{\mathbf{h}} -\\
&- 2\alpha\sqrt{P}\text{Re}\left\{zd\hat{\mathbf{h}}^H\hat{\mathbf{h}}\right\} + 1 = \\
&= z^2\alpha^2Pd^2\hat{\mathbf{h}}^H\hat{\mathbf{h}}\hat{\mathbf{h}}^H\hat{\mathbf{h}} + z^2\alpha^2Pq\hat{\mathbf{h}}^H\hat{\mathbf{h}} + z^2\sigma^2\hat{\mathbf{h}}^H\hat{\mathbf{h}} -\\
&- 2\alpha\sqrt{P}\text{Re}\left\{zd\|\hat{\mathbf{h}}\|^2\right\} + 1 = \\
&= z^2\alpha^2Pd^2\|\hat{\mathbf{h}}\|^2\|\hat{\mathbf{h}}\|^2 + z^2\alpha^2Pq\|\hat{\mathbf{h}}\|^2 + z^2\sigma^2\|\hat{\mathbf{h}}\|^2 -\\
&- 2\alpha\sqrt{P}zd\|\hat{\mathbf{h}}\|^2 + 1 = \\
&= z^2\|\hat{\mathbf{h}}\|^2\left(\alpha^2Pd^2\|\hat{\mathbf{h}}\|^2 + \alpha^2Pq + \sigma^2\right) -\\
&- 2\alpha\sqrt{P}zd\|\hat{\mathbf{h}}\|^2 + 1\ ,
\end{aligned}$$

which is equivalent with (11). ■

Proof of Theorem 1: To shorten the notation we introduce $Y = \|\hat{\mathbf{h}}\|^2$, $p = \alpha^2P$ and $b = qp + \sigma^2$ and rewrite (11) with these notations.

$$\begin{aligned}
&\text{MSE}(\hat{\mathbf{h}}) = \qquad (14)\\
&= d^2p\frac{pY^2}{(\sigma^2+pY)^2} + b\frac{pY}{(\sigma^2+pY)^2} - 2d\frac{pY}{\sigma^2+pY} + 1.
\end{aligned}$$

Next, we observe that MSE($\hat{\mathbf{h}}$) depends on $\hat{\mathbf{h}}$ only through $\|\hat{\mathbf{h}}\|^2$ and $\|\hat{\mathbf{h}}\|^4$. Since $\hat{\mathbf{h}}$ is a complex normal random vector (3) with covariance matrix $\mathbf{R} = r\mathbf{I}$ and $r = \varrho + \frac{\sigma^2}{P^p\alpha^2}$, it follows that $Y = \|\hat{\mathbf{h}}\|^2$ follows the Gamma($N_r, 1/r$) distribution (the sum of N_r independent r.v. which are exponentially distributed with parameter $1/r$) with probability density function:

$$f_Y(x) = \frac{r^{-N_r}x^{N_r-1}e^{-x/r}}{(N_r-1)!} \quad x>0. \qquad (15)$$

We will make use of the following integrals:

$$\begin{aligned}
&\int_{x=0}^{\infty}\frac{r^{-N_r}x^{N_r-1}e^{-x/r}}{(N_r-1)!}\cdot\frac{px^2}{(a+px)^2}dx = \\
&= \frac{N_r\left(-pr + e^{\frac{a}{pr}}\left(a+(1+N_r)pr\right)E_{in}\left(1+N_r,\frac{a}{pr}\right)\right)}{p^2r};
\end{aligned} \qquad (16)$$

$$\int_{x=0}^{\infty} \frac{r^{-N_r} x^{N_r-1} e^{-x/r}}{(N_r-1)!} \cdot \frac{px}{(a+px)^2} dx = \\ = \frac{-pr + e^{\frac{a}{pr}}\left(a + N_r pr\right) E_{in}\left(N_r, \frac{a}{pr}\right)}{p^2 r^2}; \qquad (17)$$

and

$$\int_{x=0}^{\infty} \frac{r^{-N_r} x^{N_r-1} e^{-x/r}}{(N_r-1)!} \cdot \frac{px}{(a+px)} dx = \\ = e^{\frac{a}{pr}} N_r E_{in}\left(1+N_r, \frac{a}{pr}\right), \qquad (18)$$

where $E_{in}(n,z) \triangleq \int_1^{\infty} e^{-zt}/t^n \; dt$ is a standard exponential integral function. Theorem 1 follows from averaging (14) according to the density function (15) and using (16)-(18):

$$\begin{aligned} \mathcal{E}\{MSE\} &= d^2 p \int_{x=0}^{\infty} \frac{px^2}{(a+px)^2} f_Y(x) dx + \\ &+ b \int_{x=0}^{\infty} \frac{px}{(a+px)^2} f_Y(x) dx - \\ &- 2d \int_{x=0}^{\infty} \frac{px}{a+px} f_Y(x) dx + 1 = \\ &= d^2 p \cdot N_r \left(e^{\frac{a}{pr}} \left(a + (1+N_r) pr \right) \cdot \right. \\ &\cdot \left. E_{in}\left(1+N_r, \frac{a}{pr}\right) - pr \right) \frac{1}{p^2 r} + \\ &+ b \cdot \left(e^{\frac{a}{pr}} \left(a + N_r pr \right) \cdot \right. \\ &\cdot \left. E_{in}\left(N_r, \frac{a}{pr}\right) - pr \right) \frac{1}{p^2 r^2} - \\ &- 2d \cdot \left(e^{\frac{a}{pr}} N_r E_{in}\left(1+N_r, \frac{a}{pr}\right) \right) + 1 = \\ &= d^2 N_r \frac{1}{pr} \left(e^{\frac{a}{pr}} a E_{in}\left(1+N_r, \frac{a}{pr}\right) + \right. \\ &+ \left. e^{\frac{a}{pr}} (1+N_r) pr E_{in}\left(1+N_r, \frac{a}{pr}\right) - pr \right) + \\ &+ b \frac{1}{p^2 r^2} \left(e^{\frac{a}{pr}} a E_{in}\left(N_r, \frac{a}{pr}\right) + \right. \\ &+ \left. e^{\frac{a}{pr}} N_r pr E_{in}\left(N_r, \frac{a}{pr}\right) - pr \right) - \\ &- 2d \cdot \left(e^{\frac{a}{pr}} N_r E_{in}\left(1+N_r, \frac{a}{pr}\right) \right) + 1 = \\ &= d^2 N_r \left(\mathcal{G}(a, 1+N_r) + \right. \\ &+ \left. pr \mathcal{G}(1+N_r, 1+N_r) - 1 \right) + \\ &+ \frac{b}{pr} \left(\mathcal{G}(a, N_r) + pr \mathcal{G}(N_r, N_r) - 1 \right) - \\ &- 2d \cdot \left(pr \mathcal{G}(N_r, 1+N_r) \right) + 1; \end{aligned}$$

where

$$\mathcal{G}(x,y) \triangleq \frac{1}{pr} e^{\frac{a}{pr}} x E_{in}\left(y, \frac{a}{pr}\right).$$

■

REFERENCES

[1] T. Kim and J. G. Andrews, "Optimal Pilot-to-Data Power Ratio for MIMO-OFDM," *IEEE Globecom*, 2005.

[2] ——, "Balancing Pilot and Data Power for Adaptive MIMO-OFDM Systems," *IEEE Globecom*, 2005.

[3] E. Golovins and N. Ventura, "Optimal Training for the SM-MIMO-OFDM Systems with MMSE Channel Estimation," *6th Annual Communication Networks and Services Research Conference*, pp. 470–477, 2008.

[4] V. K. V. Gottumukkala and H. Minn, "Capacity Analysis and Pilot-Data Power Allocation for MIMO-OFDM With Transmitter and Receiver IQ Inbalances and Residual Carrier Frequency Offset," *IEEE Trans. Veh. Techn.*, pp. 553–565, 2012.

[5] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Comm.*, vol. 9, no. 11, pp. 3590–3600, 2010.

[6] P. Soldati, M. Johansson, G. Fodor, and S. Sorrentino, "On Pilot Dimensioning in Multicell Single Input Multiple Output Systems," *IEEE Workshop on Broadband Wireless Access*, Houston, TX, USA,2011.

[7] S. M. Kay, *Fundamentals of Statistical Signal Processing, Vol. I: Estimation Theory*. Prentice Hall, 1993, no. ISBN: 0133457117.

[8] R. Gallager, "Circularly-Symmetric Gaussian Complex Vectors," *http://www.rle.mit.edu/rgallager/documents/CircSymGauss.pdf*, 2008.